ε KÚ <MASK>: INTEGRATING YORÙBÁ CULTURAL GREETINGS INTO MACHINE TRANSLATION

Idris Akinade[1,*], **Jesujoba O. Alabi**[2,*], **David I. Adelani**[3], **Clement Odoje**[1], **and Dietrich Klakow**[2]
[1] Department of Linguistics and African Languages, University of Ibadan
[2] Spoken Language Systems (LSV), Saarland University, Saarland Informatics Campus, Germany
[3] University College London
d.adelani@ucl.ac.uk
{akinadeidris,lekeclement2}@gmail.com
{jalabi,dklakow}@lsv.uni-saarland.de

ABSTRACT

This paper investigates the performance of massively multilingual neural machine translation (NMT) systems in translating Yorùbá greetings (ε kú <mask>[1]), which are a big part of Yorùbá language and culture, into English. To evaluate these models, we present IkiniYorùbá, a Yorùbá-English translation dataset containing some Yorùbá greetings, and sample use cases. We analysed the performance of different multilingual NMT systems including Google Translate and NLLB and show that these models struggle to accurately translate Yorùbá greetings into English. In addition, we trained a Yorùbá-English model by finetuning an existing NMT model on the training split of IkiniYorùbá and this achieved better performance when compared to the pre-trained multilingual NMT models, although they were trained on a large volume of data.

1 INTRODUCTION

In recent years, multilingual neural machine translation (NMT) models have shown remarkable improvement in translating both high and low-resource languages and have become widely used in various applications (Kudugunta et al., 2019; Aharoni et al., 2019; NLLB Team et al., 2022; Bapna et al., 2022). Despite this progress, NMT models still struggle to accurately translate idiomatic expressions (Fadaee et al., 2018; Baziotis et al., 2022), cultural concepts such as proverbs (Alkhresheh & AlMaaytah, 2018; Adelani et al., 2021), and common greetings, particularly in African languages like Yorùbá– a west African language, which has a rich cultural heritage.

Table 1 illustrates a Yorùbá sentence containing frequently used greeting phrases by the Yorùbá people, and the corresponding translations generated from three multilingual NMT systems, which are: Meta's NLLB (NLLB Team et al., 2022), Google Translate[2], and our own model.

Source: Ẹ kú ojúmọ́, ẹ sì kú déédé àsìkò yìí.
Target: Good morning and compliment for this period.
NLLB: You have died, and you have died to this hour.
Google Translate: Die every day, and die at this time.
Our Model: Good morning and compliment for this time.

Table 1: Translation outputs of 3 different NMT models

An examination of NLLB and Google Translate's model outputs reveals that they all fail to produce accurate translations for the input sentence. One possible explanation for this is the lack of sufficient

*Equal contribution.

[1]For simplicity of notation in the title, we make use of ε – the Beninese Yorùbá letter representation of Ẹ (which is used in Nigeria), and <mask> provides the context of greeting.

[2]https://translate.google.com/ evaluated on 23rd January 2023

training data including these types of greetings, even though they were trained on a large volume of multilingual data. Furthermore, *kú*, a common word in these kinds of greetings, has two main interpretations that could mean either death or a compliment, depending on the context. Similarly, the syntactic frame of occurrence also determines the meaning of the verb (the type of complement and adjunct), and this is due to the ambiguous nature of Yorùbá verbs. Hence, it is possible that these models were trained on data with *kú* having the meaning death.

To address this issue, this paper introduces a new dataset dubbed IkiniYorùbá, a Yorùbá-English translation dataset of popular Yorùbá greetings. We evaluate the performance of existing multilingual NMT systems on this dataset, and the results demonstrate that although current multilingual NMT systems are good at translating Yorùbá sentences into English, they struggle to accurately translate Yorùbá greetings, highlighting the need for further research in translating such cultural concepts on low-resource African languages.

2 YORÙBÁ GREETINGS

Yorùbá is a language spoken by the Yorùbá people. It is native to Nigeria, Benin and Togo with an estimate of over 40 million speakers Eberhard et al. (2020). Yorùbá makes use of 25 Latin letters excluding the Latin characters (c, q, v, x and z), and additional letters (ẹ, gb, ṣ , ọ). Yorùbá is a tonal language with three tones: low, middle and high. These tones are represented by the grave (e.g. “à ”), optional macron (e.g. “ā”) and acute (e.g. “á”) accents respectively.

Greetings are inseparable from the Yorùbá people since they are important for first impressions and are even considered to be a part of Yorùbá identity. After the abolition of the slave trade at the beginning of the 19th century, the Yorùbá indigenes who were rescued by the British warship settled in Freetown, a place in present-day Sierra Leone. People began to call them *a kú* which is a fragment attached to all forms of greetings in Yorùbá (Webster, 1966). This is because while an English speaker will say *good morning*, *happy birthday*, *merry Christmas*, and so on, the Yorùbá people would say *ẹ káàrọ̀*, *ẹ kú ọjọ́ ìbí*, and *ẹ kú ọdún kérésìmesì*. The recurrence of *ẹ kú* in their everyday conversation resulted the appellation *a kú*.

Ẹ kú has the same semantic importance as ‘good-’, ‘merry-’ and ‘happy-’ in English greetings. Without the fragment *ẹ kú* in the communication frame of greeting, the cultural knowledge shared by interlocutors will be lost.

Structurally, *ẹ kú* can be syntactically explained to have a subject-predicate relationship, rather than being a single lexeme or a prefix as claimed by most scholars. Using the paradigmatic relationship (de Saussure, 1983; Asher & Simpson, 1994) lens, *ẹ* can be replaced with any pronoun or nominal item (as described by interlocutors) with +human feature and still fit in perfectly. The +human feature is necessary because compliments are mainly for humans and *kú* requires a selection restriction to sieve out the non-human elements. Table 2 shows some of these constructions. It is equally important to note here that *ẹ kú* can also be used for supernatural beings or metaphysical beings which in this form sounds like a personification.

Greeting	Person	Meaning
O kú ìrìn	2nd person singular	Compliment for walking
A kú òde	1st person plural	Compliment for attending a party
Wọ́n kú ìjóòkó	3rd person plural	Compliment for sitting

Table 2: Some *Ẹ kú* constructions

Kú on the other hand is a transitive predicate that requires a compliment. This compliment could either be a noun that signifies time like *àárọ̀* (morning), a noun that denotes season like *ọ̀ririn*/*òtútù* (cold), a noun that points to a celebration like *kérésìmesì* (Christmas), a nominalized verb that describes an event or action like *ìjókòó* (sitting), and many more. Omitting the compliment in a greeting construction will alter the interpretation of the expression which may also change the meaning of *kú* to death.

3 Related Work

The development of machine translation systems for low-resource languages such as Yorùbá has seen a significant amount of research efforts in recent years. One major area of focus has been on curating translation datasets for these languages, which are collected using either automatic or manual methods. Examples of automatically collected datasets that include Yorùbá are JW300 (Agić & Vulić, 2019), CCMatrix (Schwenk et al., 2021), and CCAligned (El-Kishky et al., 2020). On the other hand, examples of manually translated datasets for Yoruba include MENYO-20k (Adelani et al., 2021), MAFAND-MT (Adelani et al., 2022), FLORES-101 (Goyal et al., 2022), and NTREX (Federmann et al., 2022). These datasets have been instrumental in the study, development, and improvement of machine translation systems for Yorùbá.

For example, Adelani et al. (2021) investigated how domain data quality and the use of diacritics, a crucial aspect of Yorùbá orthography, impact Yorùbá-English translations. Adebara et al. (2022) examined the effectiveness of Yorùbá-English machine translation in translating bare nouns (BN), by comparing the results obtained from using statistical machine translation methods and neural approaches. Adelani et al. (2022) investigated how to effectively leverage pre-trained models for translation of African languages including Yorùbá. Despite the attempts to create datasets and develop translation systems for Yorùbá, to the best of our knowledge, only Adelani et al. (2021) has examined a cultural aspect of Yorùbá by evaluating their models on Yorùbá proverbs, which are a significant part of Yorùbá tradition. However, this research has not looked into how these models perform on another cultural aspect which is Yorùbá greetings. Furthermore, there appear to be no prior works that have evaluated machine translation performance specifically for this aspect of the language and for other languages. Therefore, in this work, we investigate the performance of Yorùbá-English translation models on Yorùbá greetings.

4 Dataset

Greetings dataset: We introduce **IkiniYorùbá**, a Yorùbá-English translation dataset for Yorùbá greetings and their usage in various contexts, containing 960 parallel instances. The data curation process involved three key stages. Firstly, we gathered commonly used Yorùbá greetings that cover a variety of situations such as time, season, celebration, and more, as outlined in Section 2, resulting in a total of 160 Yorùbá greetings. Secondly, we created 5 different example sentences for each greeting, where the greetings are used in context, by native speakers of the language, resulting in 800 use cases in total. Lastly, we asked an expert translator to translate the seed data and the use cases into English. We split the created data into train/dev/test splits with 100/20/40 seed greeting instances. For each instance in a split, the 5 example sentences created are assigned to the same split.

Conversational dataset: For our experiments, we used the movie transcripts subset of the MENYO-20k Adelani et al. (2020) dataset, which is a human-translated English-Yorùbá dataset for movie transcripts. We selected this dataset because it consists of conversational data.

Table 3(left) illustrates the sample sentences in the IkiniYorùbá dataset (seed greeting data and sample use cases) and Movie Transcript datasets, while Table 3(right) highlights the statistics of these datasets.

5 Experiments

5.1 Experimental Setup

Greetings play a crucial role in Yorùbá culture and are widely used in daily conversations by Yorùbá people. For every action, there is a customary way of greeting or complimenting those involved using the phrase *Ẹ kú*. In this work, we compare several existing translation systems and evaluate their performance on Yorùbá greetings. We demonstrate the effectiveness of these translation systems by testing them on movie transcripts, which are conversational in nature. Below, we outline our experiments.

Yorùbá	English
IkiniYorùbá- Seed Greetings	
Ẹ kú ìfẹ́	Thanks for the love
Ọkọ̀ á rèfò	Safe ride
IkiniYorùbá- Greetings with contexts	
Ẹ kú ìfẹ́, Ire là ó má bá ara wa ṣe.	Thanks for the love, may we continue to celebrate one another.
A ó ma fojú sónà láti ríi yín, ọkọ̀ á rèfò	Looking forward to seeing you, safe ride.
Movie Transcript	
Ẹ káàsán ma.	Good afternoon ma.
Ẹ ǹlẹ̀ ṣà! Mo mọ̀ yín	Hello sir! I know you
Fẹ́mi kí ló ṣẹlẹ̀ báyìí?	Femi what is it now?
Gbogbo nnkàn á dára, a jọ wà nínú ẹ̀ ni	Everything will be fine, we're in this together

Data	Number of Sentences train	dev	test
IkiniYoruba	100/500	20/100	40/200
Movie Transcript	–	–	775

Table 3: **Left:** Sample sentence pairs from the IkiniYorùbá and the Movie Transcripts datasets. **Right:** The split of the data.

Translation Models: In this study, we evaluate the performance of three multilingual NMT systems. These systems were pre-trained on various languages, and they are Google multilingual NMT, the distilled version of Meta's NLLB (NLLB Team et al., 2022) with 600M parameters, and a publicly available M2M-100 (Fan et al., 2020) with 418M parameters fine-tuned on the MENYO-20k dataset. We generated translations for the test sets using the Google Translate web application[3], while for Meta's M2M-100 and NLLB models, we used the HuggingFace transformers[4] library.

Data preprocessing and evaluation: To standardize the format of the two parallel datasets, we converted the Yorùbá texts in the dataset to Unicode Normalization Form Composition (NFC). And to automatically assess the performance of the models, we used BLEU (Papineni et al., 2002) score implemented in SacreBLEU[5] (Post, 2018).

5.2 Experimental results

Table 4 shows the results of evaluating the three different models on the two datasets: IkiniYorùbá test split and Movie Transcripts. The models obtained impressive performance on the Movie Transcript data with high BLEU scores but poorly on the IkiniYorùbá data with significantly lower scores. This highlights their inability to translate Yorùbá cultural content such as greetings. The best-performing model, M2M-100, had a BLEU score of 34.70 on Movie Transcript data as it was trained on this same data by its authors. However, it had a score of 4.3 on greetings data. The second-best model, Google Translate, was 3.65 points below the best model on Movie Transcript. It performed better on greetings data with a score of 9.47, though still lower compared to its performance on Movie Transcript data.

In addition, we finetuned the M2M-100 model on IkiniYorùbá, Movie Transcripts, and a combination of both data sources and evaluated the models on the IkiniYorùbá test split. Our results show that finetuning the M2M-100 on Movie Transcripts improves the model's performance on IkiniYorùbá by 1.92 BLEU points compared to the original M2M-100. However, the best performance was achieved when the M2M-100 was finetuned on the IkiniYorùbá training split, with a BLEU score of 29.67. Finetuning the M2M-100 on the combination of both datasets did not result in any improvement. We do not evaluate the M2M-100 model finetuned on MovieTranscript data on the MovieTranscript data, as this would result in evaluating on the same data used for training.

[3] `https://translate.google.com/` evaluated on 23rd January 2023

[4] `https://github.com/huggingface/transformers`

[5] `case:mixed|eff:no| tok:13a|smooth:exp|version:2.3.1`

	yo → en BLEU		Adequacy	CCP
	Movie Transcript	**IkiniYorùbá**	**IkiniYorùbá**	
Google Translate	31.05	9.47	2.02	0.11
NLLB	27.19	5.03	1.88	0.09
M2M-100	34.70	4.33	1.73	0.05
+ IkiniYorùbá	26.05	**29.67**	**2.79**	**0.35**
+ Movie Transcript	-	6.25	-	-
+ IkiniYorùbá + Movies Transcript	-	29.49	-	-

Table 4: Performance of the models on IkiniYorùbá and Movie Transcript. The M2M-100 and NLLB models have 418M and 600M parameters respectively. CCP is Cultural Content Preservation and it indicates whether greetings/compliments within the source sentences are preserved or not in the translation outputs.

To understand the performance of individual models on the IkiniYorùbá test set, we conducted human evaluations of the translated outputs from Google Translate, NLLB, M2M-100, and M2M-100 finetuned on the IkiniYorùbá dataset. We asked three native Yorùbá speakers fluent in English to rate the 240 sentences for each system on two criteria: adequacy (on a Likert scale of 1 to 5) and cultural content preservation - CCP (binary scale of 0 or 1). Here, adequacy describes how much of the meaning of the reference translation was preserved in the MT output, and CCP indicates whether the greetings/compliments within the translation are preserved or not. The results show that the NMT systems struggle at translating Yorùbá greetings accurately, and they confirm the results of the automatic evaluation, showing that M2M-100 finetuned on IkiniYorùbá outperforms all other models. Overall, we observed that human evaluation shows moderate agreement with automatic evaluation.

5.3 Qualitative analysis of translation outputs

In Table 5, we present some translation outputs from the different models for 5 Yorùbá sentences sampled from the IkiniYorùbá test split.

Google Translate and NLLB perform well in some cases by generating translations that were similar and contextually appropriate, for instance, in the second and third examples. Google Translate gave the most similar output to the target sentence in the first example. Our model in this instance translated ‘ọdún’ (meaning ‘year’ in isolation or ‘celebration’ when it occurs alone with ẹ kú) quite independently ‘àjínde’ (meaning ‘resurrection’ in isolation). Hence, ‘resurrection celebration’ appears in the output. NLLB fails in this example but in the second example, it gives the closest contextual interpretation while our model got everything right except ‘àpèjẹ’ which is translated as ‘reception’ instead of ‘feasting’.

Our model outperforms Google Translate and NLLB in the third and fourth examples. It generated nearly identical output to the target sentence, thereby showing the preservation of both cultural content and semantic interpretation ability learned from the training data. In contrast, both Google Translate and NLLB were unsuccessful in producing the correct translation. The third example is an inquiry about well-being and it is, therefore, appropriate to use the word ‘fine’, and not ‘peace’. In the fourth example, our model also shows to have an understanding of the contextual usage of *kú* as a compliment which both Google Translate and NLLB failed to do. In addition, similar to the automatic evaluation result, our model generated better outputs when compared to M2M-100 which was the base model on which it was trained, confirming the ability of the model to learn from a few training instances even for low-resource languages such as Yorùbá Adelani et al. (2022).

However, all the models failed in the last example. The models incorporated the concept of celebration or birthday in their output, but none of them were able to produce output that was exactly or semantically equivalent to the target sentence. A mistake common to all the model output except for M2M-100, is that they tried to translate ‘Olúwadámiláre’[6] which is a name of a person and should not be translated. Hence, there is a need for more effort in solving this greetings translation task, either by creating more data or developing better approaches at translating these greetings into English.

[6]translates to: ‘the lord justifies me’, but the models still failed in this case.

6 Conclusion

In this study, we analyzed the performance of machine translation models in translating Yorùbá greetings into English. To achieve this objective, we introduced a novel dataset called IkiniYorùbá, which contains a collection of Yorùbá greetings and their respective sentence use cases. We evaluated three publicly available machine translation models on this dataset and found that, despite their ability to translate other Yorùbá texts, they failed to accurately translate Yorùbá greetings, which are a crucial aspect of Yorùbá culture. In future research, we aim to expand the IkiniYorùbá dataset by adding more profession-based greetings and exploring ways to enhance the performance of machine translation models with these data.

Acknowledgments

We would like to express our gratitude to Dr. Ezekiel Soremekun for initiating the We appreciate Dr. Ezekiel Soremekun for the initial discussion that led to this work. We are grateful for the feedback from Dr. Rachel Bawden, Vagrant Gautam and anonymous reviews from AfricaNLP and C3NLP. Moreover, we would like to thank Timileyin Adewusi, Ganiyat Afolabi, and Oluwatosin Koya who took part in the human evaluation process. Jesujoba Alabi was partially funded by the BMBF project SLIK under the Federal Ministry of Education and Research grant 01IS22015C. David Adelani acknowledges the support of DeepMind Academic Fellowship programme.

References

Ife Adebara, Muhammad Abdul-Mageed, and Miikka Silfverberg. Linguistically-motivated Yorùbá-English machine translation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pp. 5066–5075, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics. URL https://aclanthology.org/2022.coling-1.449.

David Adelani, Dana Ruiter, Jesujoba Alabi, Damilola Adebonojo, Adesina Ayeni, Mofe Adeyemi, Ayodele Esther Awokoya, and Cristina España-Bonet. The effect of domain and diacritics in Yoruba–English neural machine translation. In *Proceedings of Machine Translation Summit XVIII: Research Track*, pp. 61–75, Virtual, August 2021. Association for Machine Translation in the Americas. URL https://aclanthology.org/2021.mtsummit-research.6.

David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen Muhammad, Guyo Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ajibade, Tunde Ajayi, Yvonne Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. A few thousand translations go a long way! leveraging pre-trained models for African news translation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3053–3070, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.223. URL https://aclanthology.org/2022.naacl-main.223.

David Ifeoluwa Adelani, Dana Ruiter, Jesujoba Oluwadara Alabi, Damilola Adebonojo, Adesina Ayeni, Mofetoluwa Adeyemi, Ayodele Awokoya, and Cristina España-Bonet. Menyo-20k: A multi-domain english-yorùbá corpus for machine translation and domain adaptation. *ArXiv*, abs/2103.08647, 2020.

Željko Agić and Ivan Vulić. JW300: A wide-coverage parallel corpus for low-resource languages. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3204–3210, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1310. URL https://aclanthology.org/P19-1310.

Roee Aharoni, Melvin Johnson, and Orhan Firat. Massively multilingual neural machine translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 3874–3884, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1388. URL `https://aclanthology.org/N19-1388`.

Mohammad H. Al-khresheh and Shahab A. AlMaaytah. English proverbs into arabic through machine translation. *International Journal of Applied Linguistics and English Literature*, 2018.

R.E. Asher and J.M.Y. Simpson. *The Encyclopedia of Language and Linguistics*. Number Bd. 9 in The Encyclopedia of Language and Linguistics. Pergamon Press, 1994. ISBN 9780080359434. URL `https://books.google.de/books?id=5kdiAAAAMAAJ`.

Ankur Bapna, Isaac Caswell, Julia Kreutzer, Orhan Firat, Daan van Esch, Aditya Siddhant, Mengmeng Niu, Pallavi Nikhil Baljekar, Xavier Garcia, Wolfgang Macherey, Theresa Breiner, Vera Saldinger Axelrod, Jason Riesa, Yuan Cao, Mia Chen, Klaus Macherey, Maxim Krikun, Pidong Wang, Alexander Gutkin, Apu Shah, Yanping Huang, Zhifeng Chen, Yonghui Wu, and Macduff Richard Hughes. Building machine translation systems for the next thousand languages. Technical report, Google Research, 2022.

Christos Baziotis, Prashant Mathur, and Eva Hasler. Automatic evaluation and analysis of idioms in neural machine translation, 2022. URL `https://arxiv.org/abs/2210.04545`.

Ferdinand de Saussure. Course in general linguistics. 1983. (trans. Roy Harris).

David M. Eberhard, Gary F. Simons, and Charles D. Fennig (eds.). Ethnologue: Languages of the world. twenty-third edition., 2020. URL `http://www.ethnologue.com`.

Ahmed El-Kishky, Vishrav Chaudhary, Francisco Guzmán, and Philipp Koehn. CCAligned: A massive collection of cross-lingual web-document pairs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 5960–5969, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.480. URL `https://aclanthology.org/2020.emnlp-main.480`.

Marzieh Fadaee, Arianna Bisazza, and Christof Monz. Examining the tip of the iceberg: A data set for idiom translation. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA). URL `https://aclanthology.org/L18-1148`.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. Beyond english-centric multilingual machine translation. *arXiv preprint*, 2020.

Christian Federmann, Tom Kocmi, and Ying Xin. NTREX-128 – news test references for MT evaluation of 128 languages. In *Proceedings of the First Workshop on Scaling Up Multilingual Evaluation*, pp. 21–24, Online, nov 2022. Association for Computational Linguistics. URL `https://aclanthology.org/2022.sumeval-1.4`.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzmán, and Angela Fan. The Flores-101 evaluation benchmark for low-resource and multilingual machine translation. *Transactions of the Association for Computational Linguistics*, 10:522–538, 2022. doi: 10.1162/tacl_a_00474. URL `https://aclanthology.org/2022.tacl-1.30`.

Sneha Kudugunta, Ankur Bapna, Isaac Caswell, and Orhan Firat. Investigating multilingual NMT representations at scale. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 1565–1575, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1167. URL `https://aclanthology.org/D19-1167`.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia-Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. No language left behind: Scaling human-centered machine translation. 2022.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics. doi: 10.3115/1073083.1073135. URL `https://aclanthology.org/P02-1040`.

Matt Post. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pp. 186–191, Brussels, Belgium, October 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-6319. URL `https://aclanthology.org/W18-6319`.

Holger Schwenk, Guillaume Wenzek, Sergey Edunov, Edouard Grave, Armand Joulin, and Angela Fan. CCMatrix: Mining billions of high-quality parallel sentences on the web. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 6490–6500, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.507. URL `https://aclanthology.org/2021.acl-long.507`.

J. B. Webster. A preface to modern nigeria: The 'sierra leonians' in yoruba, 1830–1890. by jean herskovits kopytoff. *Africa*, 36(4):452–453, 1966. doi: 10.2307/1158056.

1.	**Source**	A kí àwọn kìrìsìtẹ́ni kú ọdún Àjínde.
	Target	We greet the Christians a happy Easter.
	Google T.	We wish Christians a happy Easter.
	NLLB	Celebrations are celebrated on New Year's Eve.
	M2M-100	We greeted ridiculers in the resurrection year.
	Our Model	We greet the hardworking people the resurrection celebration.
2.	**Source**	Ẹ kú àpèjẹ èyin olóyè.
	Target	Happy feasting chiefs.
	Google T.	Farewell to the party, you chiefs.
	NLLB	Enjoy the feast, you leaders.
	M2M-100	You chieftains die at the banquet.
	Our Model	Compliment for a reception chiefs.
3.	**Source**	Ẹ ǹlẹ́ o ẹ̀yin èèyàn mi, ṣe àlàáfíà ni?
	Target	Hello my people, I hope you are fine?
	Google T.	My people, is it peace?
	NLLB	Is it peace, my people?
	M2M-100	May you, my people, be at peace?
	Our Model	Hello my people, hope you are fine?
4.	**Source**	O kú àjàbọ́ ọ̀rẹ́ mi.
	Target	Compliment for escaping danger my friend.
	Google T.	You are dead my friend.
	NLLB	You sacrificed my friend.
	M2M-100	You lost my friend's womb.
	Our Model	Compliment for escaping the danger of my friend.
5.	**Source**	O kú ayẹyẹ ọjọ́ ìbí Olúwadámiláre.
	Target	Happy birthday celebration Olúwadámiláre.
	Google T.	He died celebrating the birthday of the Almighty.
	NLLB	You celebrated the Righteous One's birthday.
	M2M-100	You died on the anniversary of the birth of Olúwádámiler.
	Our Model	Compliment for today's anniversary of God's goodwill.

Table 5: Examples of MT output for different NMT models. Examples selected from the test set.